

Second-Order Asymptotic Optimality in Multisensor Sequential Change Detection

Georgios Fellouris* and Grigory Sokolov†

Abstract

A generalized multisensor sequential change detection problem is considered, in which a number of (possibly correlated) sensors monitor an environment in real time, the joint distribution of their observations is determined by a global parameter vector, and at some unknown time there is a change in an unknown subset of components of this parameter vector. In this setup, we consider the problem of detecting the time of the change as soon as possible, while controlling the rate of false alarms. We establish the second-order asymptotic optimality (with respect to Lorden’s criterion) of various generalizations of the CUSUM rule; that is, we show that their additional expected worst-case detection delay (relative to the one that could be achieved if the affected subset was known) remains bounded as the rate of false alarm goes to 0, for any possible subset of affected components. This general framework incorporates the traditional multisensor setup in which only an unknown subset of sensors is affected by the change. The latter problem has a special structure which we exploit in order to obtain feasible representations of the proposed schemes. We present the results of a simulation study where we compare the proposed schemes with scalable detection rules that are only first-order asymptotically optimal. Finally, in the special case that the change affects exactly one sensor, we consider the scheme that runs in parallel the local CUSUM rules and study the problem of specifying the local thresholds.

1 Introduction

Suppose that a decision maker sequentially collects data from multiple streams, coming for example from sensors monitoring an environment. At some unknown point in time there is an abrupt change in the system that is perceived by either all or only a subset of the deployed sensors. The problem then is to combine the sequentially acquired observations from all streams in order to detect the change as quickly as possible, while controlling the rate of false alarms below an acceptable level. In what follows, we will refer to the various streams as “sensors”, although this general framework can be applied to various setups beyond environmental monitoring, such as intruder detection in computer security [1–3], epidemic detection in bio-surveillance [4, 5].

*G. Fellouris is with the Department of Statistics, University of Illinois, Urbana-Champaign, IL, USA, e-mail: fellouri@illinois.edu.

†G. Sokolov is with the Department of Mathematical Sciences, Binghamton University, Binghamton, NY, USA, e-mail: gsokolov@binghamton.edu.

If the regime before and after the change is completely specified, one may apply a classical sequential change detection algorithm, such as the Cumulative Sum (CUSUM) rule, which was proposed by Page [6] and was shown by Moustakides [7] to be the best possible rule in Lorden's [8] minimax setup in the case of iid observations before and after the change. However, in many applications that motivate this multisensor sequential change detection problem it is reasonable to assume that there is only partial information regarding the post-change regime. A typical assumption in the literature is that the change affects the distribution of only a subset of sensors whose identity is unknown. In this context, one needs to design detection rules that account for this uncertainty, and also to quantify how much is lost relative to the ideal case that the actual affected subset is known.

The simplest multisensor detection rule is the *multichart* CUSUM, according to which each sensor runs locally the CUSUM algorithm and the fusion center stops the first time an alarm is raised by any sensor. This is a *one-shot* scheme suitable for decentralized implementation, as each possibly remote sensor needs to communicate with the fusion center at most once, an important advantage in sensor networks that are characterized by limited bandwidth and energy. Tartakovsky et al. [1,2] showed that this rule has a strong asymptotic optimality property when the change affects exactly *one* sensor (see also Section 9.2 in [9]). However, this rule is very inefficient when a large, or even a moderate, number of sensors are actually affected. This calls for detection rules that are robust, in the sense that they should perform well for a large class of scenarios regarding the subset of affected sensors.

In order to address this problem, Mei [10] proposed raising an alarm when the sum of the local CUSUM statistics exceeds a threshold, a rule to which we will refer as SUM-CUSUM. Under the assumptions of iid observations before/after the change, completely specified but arbitrary pre/post-change distributions (up to an integrability condition) and independent sensors it was shown that SUM-CUSUM is asymptotically optimal for every possible subset of affected sensors. Moreover, based on simulation experiments it was argued that SUM-CUSUM performs better than the multichart CUSUM unless a very small number of sensors is actually affected. This scheme was further refined by Mei in [11], where it was suggested to use the sum of only the top L local CUSUM statistics when at most L sensors can be affected. Liu et al. [12] and Banerjee and Veeravalli [13] considered further modifications that allow for only a subset of sensors to be employed at any given time.

A different, CUSUM-based, multisensor detection scheme was proposed by Xie and Siegmund [14]. Specifically, the proposed rule in this work was a modification of the CUSUM rule that corresponds to the case that the change occurs with probability π in each sensor, where π is the actual proportion of affected sensors. Asymptotic approximations for the operating characteristics of this detection rule were obtained in the case of independent sensors, each of which observes Gaussian iid observations whose mean may change from 0 to some unknown value. Moreover, it was argued based on simulation experiments that this detection rule is more efficient than SUM-CUSUM and this phenomenon was explained on the basis that SUM-CUSUM does not incorporate the fact that all sensors are assumed to have the same change-point.

The main motivation for our work was to propose and study multisensor detection rules that enjoy a stronger asymptotic optimality property than SUM-CUSUM and are more efficient in practice. However, we consider a more general setup that incorporates applications that are not captured by the traditional multisensor framework, such as the power outage detection problem treated by Chen et al. [15]. Specifically, we assume that there is a global parameter vector that determines the joint distribution of the (possibly correlated) data streams and that the change modifies a subset of the components of this parameter vector. The affected subset of components is unknown, but it is assumed to belong to a given class of subsets, \mathcal{P} , that reflects our prior information regarding the location of the change. For example, in the extreme case that there is no information regarding the affected subset, \mathcal{P} consists of all non-empty subsets of components.

In this generalized multisensor setup, we consider first of all the CUSUM-type rule that employs the Generalized Likelihood Ratio (GLR) statistic with respect to the unknown subset, to which we refer as GLR-CUSUM. Moreover, we consider two alternative, *mixture*-based CUSUM-type schemes. We show that all these detection rules enjoy a *second-order* asymptotic optimality property, i.e., that their additional worst-case detection delay (relative to the one that could be achieved if the affected subset was known) remains bounded as the rate of false alarms goes to 0 for any possible subset of affected components. This should be contrasted with the usual, weaker asymptotic optimality property, according to which the asymptotic relative efficiency goes to 1, and to which we refer as *first-order* asymptotic optimality.

Our general framework includes as a special case the traditional setup in which the change affects only an unknown subset of sensors, leaving the remaining ones completely unaffected. In this special case, assuming further independent sensors and an upper bound on the size of the affected subset, we obtain representations of the GLR-CUSUM (and one of the proposed mixture-based rule) which are feasible when the number of sensors is large. Moreover, we show that while the detection rule that was proposed by Xie and Siegmund [14] is also second-order asymptotically optimal for any choice of the parameter π , this parameter cannot be interpreted in this scheme as the proportion of affected sensors. Thus, we provide a theoretical explanation for the empirically observed superiority of this rule over SUM-CUSUM, which is only first-order asymptotically optimal, as well as its robustness with respect to π .

In the special case that it is known in advance that the change affects exactly *one* sensor, GLR-CUSUM reduces to the multichart CUSUM, according to which an alarm is raised when a local CUSUM statistic exceeds a local threshold. The selection of identical thresholds seems to be a default choice in the literature, although non-identical thresholds have also been considered (see, e.g. [9, p.467]). We address the problem of threshold selection in this setup and we show that seemingly reasonable threshold selections may lead to non-robust behavior, as the multichart CUSUM may fail to be even first-order asymptotically optimal. We then argue in favor of a specific selection of thresholds that equalizes the asymptotic *relative* performance loss under the various scenarios.

Let us discuss at this point the connection of our work with the literature of sequential change detection. Lorden [8] assumed (in a “single-sensor” setup) that the distribution of the observations

belongs to an exponential family and is specified after the change up to an unknown parameter. In this context, he established the first-order asymptotic optimality of a CUSUM-type rule that is based on (a truncated away from 0 version of) the GLR statistic. In the special case of detecting a shift in the mean of Gaussian observations, Siegmund and Venkatraman [16] obtained approximations for the operating characteristics of the corresponding GLR-CUSUM. Our work differs from these two papers in that in our setup there is a finite number of scenarios in the post-change regime. Nevertheless, we also use the term GLR-CUSUM to describe the CUSUM-type rule whose detection statistic is based on maximizing with respect to the affected subset, which is the unknown parameter in our setup.

We should also mention an alternative formulation of the multisensor sequential change detection problem, in which the change-points in the various sensors are in generally different and the goal is to detect the first of them. This problem was considered from a Bayesian point of view by Ragavan and Veeravalli [17] and Ludkovski [18], where a model for the change propagation was assumed. A minimax approach was considered by Hadjiliadis et al. [19] and Zhang et al. [20], where it was shown that the multichart CUSUM has a strong asymptotic optimality property with respect to an extended Lorden criterion in which the worst case is taken with respect to all individual change-points and the history of observations up to the minimum of the change-points.

Another closely related problem to the one consider in this paper is the joint change detection and isolation problem, in which one is interested in isolating the affected subset upon detecting the change. While the GLR-CUSUM provides a natural estimator for this subset, a proper treatment of this problem requires a different formulation, in which the misclassification rate is also controlled together with the rate of false alarms. For more details on the joint sequential change detection and isolation problem we refer to Nikiforov [21, 22], Oskiper and Poor [23], Tartakovsky [24].

The rest of the paper is organized as follows: in Section 2 we formulate a general multisensor sequential change detection problem and establish the second-order asymptotic optimality of various CUSUM-based detection rules. In Section 3 we restrict ourselves to the traditional setup of independent sensors where only a subset of them is affected by the change. In Section 4 we focus on the case that the change affects exactly one sensor and the design of the multichart CUSUM. In Section 5 we present the findings of a simulation study that illustrates our asymptotic results. We conclude in Section 6, whereas we present the proofs of most results in Appendix. In what follows we set $x^+ = \max\{x, 0\}$, we use $|\cdot|$ to denote set cardinality, and we write $x \sim y$ when $x/y \rightarrow 1$, $x \simeq y$ when $x - y \rightarrow 0$.

2 A general multisensor sequential quickest detection problem

2.1 Problem formulation

Let K be the number of sensors (streams) and X_t^k the observation in sensor k at time t , where $1 \leq k \leq K$ and $t \in \mathbb{N} = \{1, 2, \dots\}$. We assume that the random vectors $\{X_t \equiv (X_t^1, \dots, X_t^K)\}_{t \in \mathbb{N}}$ are independent over time, but we allow observations from different sensors at the same time instant to

be dependent. Moreover, we assume that the distribution of the sequence $\{X_t\}_{t \in \mathbb{N}}$ changes at some unknown deterministic time $\nu \in \mathbb{N} \cup \{0\}$. To be more specific, let $f(\cdot|\theta)$ be a family of densities with respect to some dominating σ -finite measure, where θ is a N -dimensional parameter vector and N not necessarily equal to K . Then, we assume that $X_t \sim f(\cdot|\theta_t)$, where $\theta_t := (\theta_t^1, \dots, \theta_t^N)$ is given by

$$\begin{aligned} \theta_t^k &= 0 \quad \forall 1 \leq k \leq N, \quad t \leq \nu, \\ \theta_t^k &= \begin{cases} 1, & k \in \mathcal{A}, \\ 0, & k \notin \mathcal{A}, \end{cases} \quad t > \nu, \end{aligned} \tag{1}$$

and $\mathcal{A} \subset \{1, \dots, N\}$ is a subset of components of θ_t . Note that the change may affect the observations in all sensors and that while the pre-change distribution is completely specified, the post-change distribution is specified up to subset \mathcal{A} which is typically unknown. However, we will assume that the true affected subset belongs to a given class \mathcal{P} of subsets of $\{1, \dots, N\}$. For example, if it is known that the change will affect exactly (resp. at most) L components of θ , then $\mathcal{P} = \mathcal{P}_L$ (resp. $\mathcal{P} = \overline{\mathcal{P}}_L$), where

$$\begin{aligned} \mathcal{P}_L &:= \{\mathcal{A} : |\mathcal{A}| = L\}, \\ \overline{\mathcal{P}}_L &:= \{\mathcal{A} : 1 \leq |\mathcal{A}| \leq L\}, \end{aligned} \tag{2}$$

$|\mathcal{A}|$ being the cardinality of subset \mathcal{A} and $1 \leq L \leq N$. Note that $\overline{\mathcal{P}}_N$ corresponds to the case of complete ignorance regarding the affected subset. Moreover, any class \mathcal{P} is a subset of a class of the form $\overline{\mathcal{P}}_L$ for some $1 \leq L \leq N$. However, in what follows we do not make any assumption about \mathcal{P} unless this is explicitly stated.

Let P_∞ and E_∞ the probability measure and the expectation when the change never occurs ($\nu = \infty$) and $P_\nu^{\mathcal{A}}$ and $E_\nu^{\mathcal{A}}$ the probability measure and the expectation when the change occurs at time ν only in subset \mathcal{A} . A sequential change detection rule is an $\{\mathcal{F}_t\}$ -stopping time, T , at which the fusion center raises an alarm declaring that the change has occurred. Here, \mathcal{F}_t the σ -algebra generated by all observations up to time t , i.e., $\mathcal{F}_t := \sigma(X_s; 1 \leq s \leq t)$. Following Lorden's approach [8] we measure the performance of a detection rule T *when the change occurs in subset \mathcal{A}* with the worst (with respect to the change point) conditional expected delay given the worst possible history of observations up to the change point:

$$\mathcal{J}_{\mathcal{A}}[T] := \sup_{\nu \geq 0} \text{ess sup } E_\nu^{\mathcal{A}}[(T - \nu)^+ | \mathcal{F}_\nu].$$

Moreover, we control the rate of false alarms below an acceptable, user-specified level $\gamma \geq 1$, thus, we restrict ourselves to sequential detection rules in the class $\mathcal{C}_\gamma = \{T : E_\infty[T] \geq \gamma\}$. We will be interested in sequential detection rules that attain the optimal performance $\inf_{T \in \mathcal{C}_\gamma} \mathcal{J}_{\mathcal{A}}[T]$ for every $\mathcal{A} \in \mathcal{P}$ asymptotically as $\gamma \rightarrow \infty$.

Definition 1. Suppose that the detection rule T^* can be designed so that $T^* \in \mathcal{C}_\gamma$ for any given $\gamma \geq 1$. We will say that T^* is asymptotically optimal of

(i) first-order, if for every $\mathcal{A} \in \mathcal{P}$ we have as $\gamma \rightarrow \infty$

$$\mathcal{J}_{\mathcal{A}}[T^*] \sim \inf_{T \in \mathcal{C}_{\gamma}} \mathcal{J}_{\mathcal{A}}[T],$$

(ii) second-order, if for every $\mathcal{A} \in \mathcal{P}$ we have

$$\mathcal{J}_{\mathcal{A}}[T^*] - \inf_{T \in \mathcal{C}_{\gamma}} \mathcal{J}_{\mathcal{A}}[T] = \mathcal{O}(1),$$

where $\mathcal{O}(1)$ is a bounded term as $\gamma \rightarrow \infty$.

Remark: If T^* is first-order asymptotically optimal, then it is possible that for some $\mathcal{A} \in \mathcal{P}$ we have $\mathcal{J}_{\mathcal{A}}[T^*] - \inf_{T \in \mathcal{C}_{\gamma}} \mathcal{J}_{\mathcal{A}}[T] \rightarrow \infty$, since both terms go to infinity as $\gamma \rightarrow \infty$. In such a case, the *additional* (worst-case) expected number of observations that T^* requires (relative to the corresponding optimal detection delay that would be achievable if the actual affected subset was known) is unbounded as the rate of false alarms goes to 0. This motivates the stronger notion of second-order asymptotic optimality.

It is the main goal of this work to propose detection rules that are second-order asymptotically optimal. In order to do so, we need to characterize the optimal performance that is attained by the oracle detection rule.

2.2 The oracle rule

For any given subset \mathcal{A} , the solution to the constrained optimization problem $\inf_{T \in \mathcal{C}_{\gamma}} \mathcal{J}_{\mathcal{A}}[T]$ is well known. In order to describe it, let us set

$$\Lambda_t^{\mathcal{A}} := \frac{d\mathbf{P}_0^{\mathcal{A}}}{d\mathbf{P}_{\infty}}(\mathcal{F}_t) \quad \text{and} \quad Z_t^{\mathcal{A}} := \log \Lambda_t^{\mathcal{A}} \quad (3)$$

and note that due to the assumption of independence over time we have

$$\begin{aligned} \Lambda_t^{\mathcal{A}} &= \Lambda_{t-1}^{\mathcal{A}} \cdot \exp\{\ell_t^{\mathcal{A}}\}, \quad \Lambda_0^{\mathcal{A}} := 1, \\ Z_t^{\mathcal{A}} &= Z_{t-1}^{\mathcal{A}} + \ell_t^{\mathcal{A}}, \quad Z_0^{\mathcal{A}} := 0, \end{aligned}$$

where $\ell_t^{\mathcal{A}}$ is the log-likelihood ratio

$$\ell_t^{\mathcal{A}} := \log \frac{d\mathbf{P}_0^{\mathcal{A}}}{d\mathbf{P}_{\infty}}(X_t). \quad (4)$$

The CUSUM rule for detecting a change in subset \mathcal{A} is

$$S_b^{\mathcal{A}} = \inf \left\{ t \in \mathbb{N} : \tilde{Y}_t^{\mathcal{A}} \geq b \right\}, \quad (5)$$

where $b \in \mathbb{R}$ is a deterministic threshold and $\tilde{Y}^{\mathcal{A}}$ is the so-called CUSUM statistic

$$\tilde{Y}_t^{\mathcal{A}} := Z_t^{\mathcal{A}} - \min_{0 \leq s < t} Z_s^{\mathcal{A}}, \quad t \in \mathbb{N}, \quad (6)$$

which may be equivalently defined through the following recursion

$$\tilde{Y}_t^{\mathcal{A}} = \max\{\tilde{Y}_{t-1}^{\mathcal{A}}, 0\} + \ell_t^{\mathcal{A}}, \quad \tilde{Y}_0^{\mathcal{A}} := 0. \quad (7)$$

Moustakides [7] showed that, for any given $\gamma \geq 1$, $S_b^{\mathcal{A}}$ optimizes Lorden's criterion, $\mathcal{J}_{\mathcal{A}}$, in the class \mathcal{C}_{γ} , when threshold b is chosen so that $\mathbb{E}_{\infty}[S_b^{\mathcal{A}}] = \gamma$. Earlier, Lorden [8] had established the *first-order* asymptotic optimality of the CUSUM rule by showing that

$$\inf_{T \in \mathcal{C}_{\gamma}} \mathcal{J}_{\mathcal{A}}[T] \sim \frac{\log \gamma}{I_{\mathcal{A}}}, \quad I_{\mathcal{A}} := \mathbb{E}_0^{\mathcal{A}}[\ell_1^{\mathcal{A}}], \quad (8)$$

under the assumption that $\ell_1^{\mathcal{A}}$ is $\mathbb{P}_0^{\mathcal{A}}, \mathbb{P}_{\infty}$ -integrable. However, the exact optimality of the CUSUM rule can be used to obtain a second-order characterization of the optimal performance. Indeed, if additionally we have $\mathbb{E}_0^{\mathcal{A}}[(\ell_1^{\mathcal{A}})^2] < \infty$, it is well known (see, e.g., [9], [25]) that

$$\mathbb{E}_0^{\mathcal{A}}[S_b^{\mathcal{A}}] = \frac{b}{I_{\mathcal{A}}} + \Theta(1),$$

where $\Theta(1)$ is a term that is bounded from above and below as $b \rightarrow \infty$. Moreover, from [26] it follows that $\mathbb{E}_{\infty}[S_b^{\mathcal{A}}] \sim e^b/v_{\mathcal{A}}$ as $b \rightarrow \infty$, where $v_{\mathcal{A}}$ is a model-dependent, renewal-theoretic constant. Therefore, we conclude that

$$\inf_{T \in \mathcal{C}_{\gamma}} \mathcal{J}_{\mathcal{A}}[T] = \frac{\log \gamma}{I_{\mathcal{A}}} + \Theta(1).$$

We summarize these characterizations of first and second order asymptotic optimality in the following Lemma.

Lemma 1. (i) Suppose that $\ell_1^{\mathcal{A}}$ is integrable with respect to \mathbb{P}_{∞} and $\mathbb{P}_0^{\mathcal{A}}$ for every $\mathcal{A} \in \mathcal{P}$ and that, for any given $\gamma \geq 1$, a detection rule T^* can be designed so that $T^* \in \mathcal{C}_{\gamma}$. If

$$\mathcal{J}_{\mathcal{A}}[T^*] \leq \frac{\log \gamma}{I_{\mathcal{A}}}(1 + o(1)), \quad \forall \mathcal{A} \in \mathcal{P},$$

then T^* is first-order asymptotically optimal with respect to class \mathcal{P} .

(ii) Suppose additionally that $\ell_1^{\mathcal{A}}$ is square-integrable with respect to $\mathbb{P}_0^{\mathcal{A}}$ for every $\mathcal{A} \in \mathcal{P}$. If

$$\mathcal{J}_{\mathcal{A}}[T^*] \leq \frac{\log \gamma}{I_{\mathcal{A}}} + \mathcal{O}(1), \quad \forall \mathcal{A} \in \mathcal{P},$$

then T^* is second-order asymptotically optimal with respect to class \mathcal{P} .

Remark: It is useful to stress that the CUSUM statistic is often defined in the literature as

$$Y_t^{\mathcal{A}} := Z_t^{\mathcal{A}} - \min_{0 \leq s \leq t} Z_s^{\mathcal{A}}, \quad (9)$$

or equivalently through the following recursion

$$Y_t^{\mathcal{A}} = \max\{Y_{t-1}^{\mathcal{A}} + \ell_t^{\mathcal{A}}, 0\}, \quad Y_0^{\mathcal{A}} := 0.$$

Thus, $Y_t^{\mathcal{A}}$ differs from $\tilde{Y}_t^{\mathcal{A}}$ only when the latter is negative, and the two statistics lead to the same stopping time when the threshold b is positive. However, with the non-negative version of the CUSUM statistic it is not in general possible to satisfy the false alarm constraint for any $\gamma > 1$. Since in this work we focus on asymptotic optimality results, we will use both versions of the CUSUM statistic, depending on which one is more technically convenient each time.

2.3 Generalized and Mixture based CUSUM rules

We now construct various multichannel detection rules that will be shown to be second-order asymptotically optimal with respect to any given class \mathcal{P} . In order to do so, suppose for the moment that the affected subset is $\mathcal{A} \in \mathcal{P}$. At some time $t > \nu$, the likelihood ratio of $\mathbf{P}_\nu^{\mathcal{A}}$ versus \mathbf{P}_∞ is given by

$$\frac{\Lambda_t^{\mathcal{A}}}{\Lambda_\nu^{\mathcal{A}}} = \frac{d\mathbf{P}_\nu^{\mathcal{A}}}{d\mathbf{P}_\infty}(\mathcal{F}_t) = \prod_{s=\nu+1}^t \exp(\ell_s^{\mathcal{A}}),$$

and the corresponding log-likelihood ratio takes the form

$$Z_t^{\mathcal{A}} - Z_\nu^{\mathcal{A}} = \log \frac{\Lambda_t^{\mathcal{A}}}{\Lambda_\nu^{\mathcal{A}}} = \sum_{s=\nu+1}^t \ell_s^{\mathcal{A}},$$

where $\Lambda_t^{\mathcal{A}}$, $Z_t^{\mathcal{A}}$ and $\ell_t^{\mathcal{A}}$ are defined in (3) and (4), respectively. Maximizing with respect to both the unknown change-point, ν , and the unknown subset, \mathcal{A} , suggests raising an alarm when the statistic

$$\max_{0 \leq s < t} \max_{\mathcal{A} \in \mathcal{P}} (Z_t^{\mathcal{A}} - Z_s^{\mathcal{A}})$$

becomes larger than some threshold b . Interchanging the two maximizations reveals that this detection statistic can be equivalently expressed as

$$\max_{\mathcal{A} \in \mathcal{P}} \max_{0 \leq s < t} (Z_t^{\mathcal{A}} - Z_s^{\mathcal{A}}) = \max_{\mathcal{A} \in \mathcal{P}} \tilde{Y}_t^{\mathcal{A}},$$

where $\tilde{Y}_t^{\mathcal{A}}$ is defined in (6). Therefore, this rule stops the first time t there is a subset \mathcal{A} whose corresponding CUSUM statistic $\tilde{Y}_t^{\mathcal{A}}$ exceeds threshold b . If we only consider positive thresholds, we may equivalently consider the detection statistic $\max_{\mathcal{A} \in \mathcal{P}} Y_t^{\mathcal{A}}$, where $Y^{\mathcal{A}}$ is defined in (9) (see the remark in the end of the previous subsection).

We can generalize this scheme by allowing the thresholds that correspond to the various CUSUM statistics to differ. Thus, if $\{b_{\mathcal{A}}, \mathcal{A} \in \mathcal{P}\}$ is a family of positive, constant thresholds, we obtain the following detection rule, to which we refer as GLR-CUSUM:

$$\min_{\mathcal{A} \in \mathcal{P}} S_{b_{\mathcal{A}}}^{\mathcal{A}} = \inf \left\{ t \in \mathbb{N} : \max_{\mathcal{A} \in \mathcal{P}} (Y_t^{\mathcal{A}} - b_{\mathcal{A}}) \geq 0 \right\}, \quad (10)$$

(Recall that $S_b^{\mathcal{A}}$ is the CUSUM rule defined in (5)). We will show in Proposition 3 that some intuitively reasonable threshold specifications, such as selecting each threshold $b_{\mathcal{A}}$ proportionally to the Kullback-Leibler divergence $I_{\mathcal{A}}$, may fail to guarantee even the first-order asymptotic optimality of GLR-CUSUM. For this reason, in what follows we restrict our attention to thresholds of the form

$$b_{\mathcal{A}} = b - \log p_{\mathcal{A}},$$

where b is a positive threshold that is determined by the false alarm constraint and depends on γ , whereas $\{p_{\mathcal{A}}, \mathcal{A} \in \mathcal{P}\}$ are constants (*weights*) that do *not* depend on γ and satisfy

$$p_{\mathcal{A}} > 0 \quad \forall \mathcal{A} \in \mathcal{P} \quad \text{and} \quad \sum_{\mathcal{A} \in \mathcal{P}} p_{\mathcal{A}} = 1.$$

With this threshold specification the GLR-CUSUM in (10) takes the form

$$S_b := \inf \left\{ t \in \mathbb{N} : \max_{\mathcal{A} \in \mathcal{P}} (Y_t^{\mathcal{A}} + \log p_{\mathcal{A}}) \geq b \right\} \quad (11)$$

and may be equivalently expressed as follows

$$\begin{aligned} S_b &= \inf \left\{ t \in \mathbb{N} : \max_{0 \leq s \leq t} \max_{\mathcal{A} \in \mathcal{P}} (Z_t^{\mathcal{A}} - Z_s^{\mathcal{A}} + \log p_{\mathcal{A}}) \geq b \right\} \\ &= \inf \left\{ t \in \mathbb{N} : \max_{0 \leq s \leq t} \max_{\mathcal{A} \in \mathcal{P}} p_{\mathcal{A}} \exp(Z_t^{\mathcal{A}} - Z_s^{\mathcal{A}}) \geq e^b \right\}. \end{aligned} \quad (12)$$

When the $p_{\mathcal{A}}$'s are identical, we will refer to S_b as the *unweighted* GLR-CUSUM. However, the introduction of weights allows us to present the GLR-CUSUM together with competitive *mixture-based* CUSUM tests. Specifically, if we do not maximize but average with respect to the affected subset in (12), we obtain the following stopping time:

$$\bar{S}_b := \inf \left\{ t \in \mathbb{N} : \max_{0 \leq s \leq t} \sum_{\mathcal{A} \in \mathcal{P}} p_{\mathcal{A}} \exp(Z_t^{\mathcal{A}} - Z_s^{\mathcal{A}}) \geq e^b \right\}. \quad (13)$$

If we further interchange maximization and summation in the detection statistic in (13), we obtain the alternative detection statistic $\sum_{\mathcal{A} \in \mathcal{P}} p_{\mathcal{A}} \exp(Y_t^{\mathcal{A}})$. For technical convenience, we replace $Y^{\mathcal{A}}$ with $\tilde{Y}^{\mathcal{A}}$ and set

$$\tilde{S}_b := \inf \left\{ t \in \mathbb{N} : \sum_{\mathcal{A} \in \mathcal{P}} p_{\mathcal{A}} \exp(\tilde{Y}_t^{\mathcal{A}}) \geq e^b \right\}. \quad (14)$$

To sum up, we have introduced three distinct multichannel detection rules: the GLR-CUSUM, S_b , which can be equivalently defined by either (11) or (12), and two *mixture-based* CUSUM rules, \bar{S}_b and \tilde{S}_b , defined in (13) and (14) respectively. In the following theorem, whose proof can be found in the [Appendix](#), we establish their second-order asymptotic optimality property with respect to an arbitrary class \mathcal{P} .

Theorem 1. *For an arbitrary class \mathcal{P} ,*

- (i) *if $b = \log \gamma$, then $S_b, \bar{S}_b, \tilde{S}_b \in \mathcal{C}_{\gamma}$;*
- (ii) *if $\ell_1^{\mathcal{A}}$ is square-integrable with respect to $\mathbf{P}_0^{\mathcal{A}}$ for every $\mathcal{A} \in \mathcal{P}$, then S, \bar{S}, \tilde{S} are second-order asymptotically optimal with respect to \mathcal{P} .*

Although it was not needed in the proof of the previous theorem, it is important (and useful for computational purposes) to note that the worst-case scenario for all the above detection rules is when $\nu = 0$. This is the content of the following proposition, whose proof can be found in the [Appendix](#).

Proposition 1. *For any $b > 0$ and $\mathcal{A} \in \mathcal{P}$ we have*

$$i) \mathcal{J}_{\mathcal{A}}[S_b] = \mathbf{E}_0^{\mathcal{A}}[S_b], \quad ii) \mathcal{J}_{\mathcal{A}}[\bar{S}_b] = \mathbf{E}_0^{\mathcal{A}}[\bar{S}_b], \quad iii) \mathcal{J}_{\mathcal{A}}[\tilde{S}_b] = \mathbf{E}_0^{\mathcal{A}}[\tilde{S}_b].$$

2.4 Implementation

The two representations of the GLR-CUSUM, (11) and (12), suggest two possible approaches regarding its implementation. According to the first one, at each time t we compute and maximize over the CUSUM statistics, $Y_t^{\mathcal{A}}$, $\mathcal{A} \in \mathcal{P}$. This approach requires running $|\mathcal{P}|$ recursions and is convenient when the cardinality of \mathcal{P} is small. This is for example the case when only one component of the parameter vector θ is affected by the change ($\mathcal{P} = \mathcal{P}_1$), in which case $|\mathcal{P}| = N$. However, such an implementation may not be feasible for large values of N when we have complete ignorance regarding the affected subset ($\mathcal{P} = \overline{\mathcal{P}}_N$). (This approach also applies to the mixture-based CUSUM rule defined in (14), which sums over the exponents of all CUSUM statistics).

According to the second approach, at each time t we need to compute and maximize over the statistics $Z_t^{\mathcal{A}} - Z_s^{\mathcal{A}} + \log p_{\mathcal{A}}$, $0 \leq s < t$, $\mathcal{A} \in \mathcal{P}$. In the next section we will see that this computation is simplified in the special case that we have independent sensors and only a subset of which is affected by the change. Moreover, in the next lemma we show that, without any loss, we can always restrict the maximization to an adaptive time-window. Since a similar approach applies to the mixture-based CUSUM defined in (13), we state the following lemma in some generality.

Lemma 2. *Let \mathcal{P} be an arbitrary class and consider the following sequence of stopping times*

$$r_{n+1} = \inf \left\{ t > r_n : Z_t^{\mathcal{A}} = \min_{0 \leq s \leq t} Z_s^{\mathcal{A}}, \text{ for all } \mathcal{A} \in \mathcal{P} \right\}; r_0 = 0. \quad (15)$$

Then for any function $G : \mathbb{R}^{|\mathcal{P}|} \rightarrow \mathbb{R}$ that is non-decreasing in each of its arguments,

$$\begin{aligned} & \inf \left\{ t \geq 0 : \max_{0 \leq s \leq t} G((Z_t^{\mathcal{A}} - Z_s^{\mathcal{A}})_{\mathcal{A} \in \mathcal{P}}) \geq b \right\} \\ &= \inf \left\{ t \geq 0 : \max_{r(t) \leq s \leq t} G((Z_t^{\mathcal{A}} - Z_s^{\mathcal{A}})_{\mathcal{A} \in \mathcal{P}}) \geq b \right\}, \end{aligned}$$

where $(\cdot)_{\mathcal{A} \in \mathcal{P}}$ denotes a \mathcal{P} -tuple and $r(t) = \max \{r_n : r_n \leq t\}$.

Proof. Fix $t > 0$. Since G is non-decreasing in each of its arguments, it suffices to show that for every $0 \leq s \leq r(t)$ and $\mathcal{A} \in \mathcal{P}$ we have $Z_t^{\mathcal{A}} - Z_s^{\mathcal{A}} \leq Z_t^{\mathcal{A}} - Z_{r(t)}^{\mathcal{A}}$, or equivalently $Z_{r(t)}^{\mathcal{A}} \leq Z_s^{\mathcal{A}}$, which follows from the definition of $r(t)$. \square

3 A special multisensor sequential change-detection problem

3.1 A special case of the general framework

In this section, we focus on the special case that the change affects the distribution of only a subset of sensors. In order to be more specific, let us recall that at each time t we observe a random vector $X_t = (X_t^1, \dots, X_t^K)$ with joint density $X_t \sim f(\cdot|\theta_t)$, where $\theta_t = (\theta_t^1, \dots, \theta_t^N)$ is an unobserved parameter vector. Let f_k be the marginal density of the k^{th} stream, i.e., $X_t^k \sim f_k(\cdot|\theta_t)$. In what follows, we assume that the dimensions of X_t and θ_t coincide, i.e., $N = K$, and that each marginal

density $f_k(\cdot|\theta_t)$ is determined only by the k^{th} component of θ_t , θ_t^k . Specifically, we assume that

$$f_k(\cdot|\theta_t^k = 0) = h_k, \quad f_k(\cdot|\theta_t^k = 1) = g_k,$$

where h_k and g_k are completely specified densities with respect to some σ -finite measure λ_k . Note that h_k and g_k do not need to belong in the same parametric family, our only assumption is that their Kullback-Leibler information number,

$$I_k := \int \log \left(\frac{g_k}{h_k} \right) g_k d\lambda_k, \quad (16)$$

is positive and finite. Therefore, in this setup we have a change in the marginal distributions of a subset of sensors, while the remaining ones remain completely unaffected, and the change detection problem (1) takes the form

$$\begin{aligned} X_t^k &\sim h_k, \quad \forall 1 \leq k \leq K, \quad t \leq \nu, \\ X_t^k &\sim \begin{cases} g_k, & k \in \mathcal{A}, \\ h_k, & k \notin \mathcal{A}, \end{cases} \quad t > \nu, \end{aligned} \quad (17)$$

where $\mathcal{A} \subset \{1, \dots, K\}$ is a subset of sensors that belongs to some class \mathcal{P} . When in particular $\mathcal{P} = \mathcal{P}_L$ (resp. $\mathcal{P} = \overline{\mathcal{P}}_L$), the change affects exactly (resp. at most) L sensors, where the classes \mathcal{P}_L and $\overline{\mathcal{P}}_L$ are defined in (2).

In the literature of the multisensor quickest detection problem it is typically assumed that observations from different streams are independent. However, this is not necessary for the results of the previous section to hold. This is illustrated by the following example.

Example: Correlated normal streams

Let Σ be an invertible covariance matrix of dimension K with diagonal $(\sigma_1^2, \dots, \sigma_K^2)$ and let μ_1, \dots, μ_K be non-zero constants. For every non-empty subset $\mathcal{A} \subset \{1, \dots, K\}$ we define the K -dimensional vector $\mu_{\mathcal{A}} := (\mu_1^{\mathcal{A}}, \dots, \mu_K^{\mathcal{A}})$ such that

$$\mu_k^{\mathcal{A}} := \begin{cases} \mu_k, & k \in \mathcal{A}, \\ 0, & k \notin \mathcal{A}, \end{cases}$$

and we assume that under $\mathbb{P}_{\nu}^{\mathcal{A}}$ we have

$$X_t \sim \begin{cases} \mathcal{N}(0, \Sigma), & t \leq \nu, \\ \mathcal{N}(\mu_{\mathcal{A}}, \Sigma), & t > \nu. \end{cases}$$

Then, this is a special case of (17) with $h_k = \mathcal{N}(0, \sigma_k^2)$ and $g_k = \mathcal{N}(\mu_k, \sigma_k^2)$, for which

$$\ell_t^{\mathcal{A}} = \theta_{\mathcal{A}} \cdot X_t - \frac{1}{2} \theta_{\mathcal{A}} \cdot \mu_{\mathcal{A}}, \quad \theta_{\mathcal{A}} := \Sigma^{-1} \mu_{\mathcal{A}},$$

where $\ell_t^{\mathcal{A}}$ is the log-likelihood ratio defined in (4).

3.2 The case of independent sensors

We now restrict ourselves to the case of independent sensors. Specifically, we assume that the local filtrations $\{\mathcal{F}_t^1\}, \dots, \{\mathcal{F}_t^K\}$ are independent, where $\mathcal{F}_t^k := \sigma(X_s^k, 1 \leq s \leq t)$, $1 \leq k \leq K$. Under this assumption, the log-likelihood ratio statistic ℓ_t^A , defined in (4), admits the following decomposition

$$\ell_t^A = \sum_{k \in \mathcal{A}} \ell_t^k, \quad \ell_t^k := \log \frac{g_k}{h_k}(X_t^k), \quad (18)$$

which implies that the likelihood and log-likelihood ratio statistics, Λ_t^A and Z_t^A , defined in (3), take the form

$$\begin{aligned} \Lambda_t^A &= \prod_{k \in \mathcal{A}} \Lambda_t^k, \quad \Lambda_t^k := \Lambda_{t-1}^k \exp\{\ell_t^k\}, \quad \Lambda_0^k = 1, \\ Z_t^A &= \sum_{k \in \mathcal{A}} Z_t^k, \quad Z_t^k := Z_{t-1}^k + \ell_t^k, \quad Z_0^k = 0. \end{aligned}$$

Moreover, we have $I_A = \sum_{k \in \mathcal{A}} I_k$, where I_A is the Kullback-Leibler information number defined in (8) and I_k the local Kullback-Leibler information number defined in (16). If we further select weights of the form

$$p_A = \frac{p^{|\mathcal{A}|}}{\sum_{B \in \mathcal{P}} p^{|\mathcal{B}|}}, \quad (19)$$

where p is some arbitrary positive parameter, then from (12) and (13) it follows that the GLR-CUSUM and the mixture-based CUSUM rule, \bar{S}_b , can be expressed as follows:

$$S_b := \inf \left\{ t \in \mathbb{N} : \max_{0 \leq s \leq t} \max_{\mathcal{A} \in \mathcal{P}} \sum_{k \in \mathcal{A}} (Z_t^k - Z_s^k + \log p) \geq b + \log(C(\mathcal{P})) \right\}, \quad (20)$$

$$\bar{S}_b := \inf \left\{ t \in \mathbb{N} : \max_{0 \leq s \leq t} \sum_{\mathcal{A} \in \mathcal{P}} p^{|\mathcal{A}|} \prod_{k \in \mathcal{A}} \exp(Z_t^k - Z_s^k) \geq e^b C(\mathcal{P}) \right\}, \quad (21)$$

where $C(\mathcal{P}) := \sum_{B \in \mathcal{P}} p^{|\mathcal{B}|}$. In the following proposition, whose proof is presented in the [Appendix](#), we show that the latter expressions can be further simplified for classes of the form \mathcal{P}_L and $\bar{\mathcal{P}}_L$. In order to do so, for any $0 \leq s \leq t$ we set

$$Z_{s:t}^k := Z_t^k - Z_s^k, \quad 1 \leq k \leq K,$$

and we introduce the corresponding order statistics $Z_{s:t}^{(1)} \geq \dots \geq Z_{s:t}^{(K)}$.

Proposition 2. *Consider the change detection problem (17) and suppose that the independence assumption (18) holds.*

(i) *The GLR-CUSUM in (20) takes the form*

$$\inf \left\{ t \in \mathbb{N} : \max_{0 \leq s \leq t} \sum_{k=1}^L Z_{s:t}^{(k)} \geq b + \log |\mathcal{P}_L| \right\} \quad \text{when } \mathcal{P} = \mathcal{P}_L, \quad (22)$$

$$\inf \left\{ t \in \mathbb{N} : \max_{0 \leq s \leq t} \sum_{k=1}^L (Z_{s:t}^{(k)} + \log p)^+ \geq b + \log \left(\sum_{k=1}^L |\mathcal{P}_k| p^k \right) \right\} \quad \text{when } \mathcal{P} = \bar{\mathcal{P}}_L. \quad (23)$$

(ii) When $\mathcal{P} = \overline{\mathcal{P}}_K$, the mixture rule \bar{S}_b in (21) takes the form

$$\hat{S}_b(\pi) := \inf \left\{ t \in \mathbb{N} : \max_{0 \leq s \leq t} \prod_{k=1}^K \left[1 - \pi + \pi \exp(Z_t^k - Z_s^k) \right] \geq \exp\{\hat{b}\} \right\}, \quad (24)$$

where $\pi = p/(1+p)$ and

$$\exp\{\hat{b}\} := (1 - \pi)^{-K} \left(\exp\{b\} \sum_{\mathcal{B} \in \mathcal{P}} p^{|\mathcal{B}|} + 1 \right).$$

Remark: From Lemma 2 we know that we can replace $\max_{0 \leq s \leq t}$ in the above detection rules by $\max_{r(t) \leq s \leq t}$, where $r(t) = \max\{n : r_n \leq t\}$ and the sequence of stopping times (r_n) are defined in (15). In the case of independent sensors that consider in this section we have $Y_t^{\mathcal{A}} \leq \sum_{k \in \mathcal{A}} Y_t^k$ for every t and \mathcal{A} , which implies that for a class of the form $\overline{\mathcal{P}}_L$ the times (r_n) can be defined as follows:

$$r_n = \inf \left\{ t > r_{n-1} : Z_t^k = \min_{0 \leq s \leq t} Z_s^k, \quad \forall 1 \leq k \leq K \right\},$$

where $r_0 := 0$. From Proposition 1 in [10] it follows that $(r_n - r_{n-1})_{n \in \mathbb{N}}$ are iid random variables with finite expectation under \mathbf{P}_∞ , which however grows exponentially in K . This implies that the above window may not be very useful for computational purposes. A more appropriate approach for practical implementation when K is large is to use instead the window $\sigma(t) := \max\{\sigma_n : \sigma_n \leq t\}$, where

$$\sigma_n := \inf \left\{ t > \sigma_{n-1} : Z_t^k < Z_{\sigma_{n-1}}^k, \quad \forall 1 \leq k \leq K \right\}; \quad \sigma_0 := 0.$$

Again, $(\sigma_n - \sigma_{n-1})_{n \in \mathbb{N}}$ is a sequence of iid random variables with finite expectation under \mathbf{P}_∞ , which however seems to grow logarithmically in K (based on empirical observations). However, we should emphasize that the resulting detection rules are not equivalent to the original ones and their asymptotic performance requires separate analysis. For similar adaptive window-limited modifications of CUSUM-type rules we refer to Yashchin [27–29].

Remark: When the weights are selected according to (19), all subsets of the same size have the same weight. It is straightforward to generalize the previous results when $p_{\mathcal{A}}$ is proportional to $\prod_{k \in \mathcal{A}} p_k$, where p_k , $1 \leq k \leq K$ are arbitrary positive parameters. However, setting $p_k = p$ offers an intuitive way of selecting the parameter π , or equivalently p , in the mixture-based CUSUM, $\hat{S}_b(\pi)$, in (24). Indeed, $\hat{S}_b(\pi)$ can be obtained by repeated application of the one-sided sequential test

$$\begin{aligned} \hat{T}_b(\pi) &:= \inf \left\{ t \in \mathbb{N} : \hat{\Lambda}_t \geq e^b \right\}, \quad \text{where} \\ \hat{\Lambda}_t &:= \prod_{k=1}^K \left(1 - \pi + \pi \Lambda_t^k \right). \end{aligned} \quad (25)$$

This is the one-sided SPRT for testing \mathbf{P}_∞ against the auxiliary probability measure

$$\hat{\mathbf{P}} = \times_{k=1}^K \left[(1 - \pi) \mathbf{P}_\infty^k + \pi \mathbf{P}_0^k \right],$$

according to which the density in sensor k is g_k with probability π and h_k with probability $1 - \pi$. This implies that the parameter π in $\hat{S}_b(\pi)$ can be interpreted as the proportion of affected sensors and suggests setting $\pi = L/K$ when we know in advance that exactly L sensors are affected ($\mathcal{P} = \mathcal{P}_L$). On the other hand, if it is known that *at most* L sensors are affected ($\mathcal{P} = \overline{\mathcal{P}}_L$), a reasonable default choice seems to be $\pi = L/(2K)$.

Remark: It is interesting to compare the detection rule \hat{S}_b , defined in (24), with the detection rule

$$\check{S}_b(\pi) := \inf \left\{ t \in \mathbb{N} : \max_{0 \leq s \leq t} \sum_{k=1}^K \log \left(1 - \pi + \pi \exp\{(Z_t^k - Z_s^k)^+\} \right) \geq b \right\}, \quad (26)$$

which was proposed by Xie and Siegmund [14]. In the following theorem, whose proof is presented in the [Appendix](#), we show that the latter is also second-order asymptotically optimal with respect to $\overline{\mathcal{P}}_K$ for any choice of π in $(0, 1]$, a result that explains the (empirically observed in [14]) robustness of this rule with respect to π .

Theorem 2. *Consider the change detection problem (17) and suppose that the independence assumption (18) holds. Consider the detection rules $\hat{S}_b(\pi)$ with some $\pi \in (0, 1)$ and $\check{S}_b(\pi)$ with some $\pi \in (0, 1]$, defined in (24) and (26), respectively. Then,*

- (i) $\hat{S}_b(\pi) \in \mathcal{C}_\gamma$ if $b = \log \gamma$ and $\check{S}_b(\pi) \in \mathcal{C}_\gamma$ if $b = \log \gamma + \log |\overline{\mathcal{P}}_K|$;
- (ii) if also ℓ_1^A is square-integrable under \mathbb{P}_0^A for every $A \in \overline{\mathcal{P}}_K$, both $\check{S}_b(\pi)$ and $\hat{S}_b(\pi)$ are second-order asymptotically optimal with respect to $\overline{\mathcal{P}}_K$.

Remark: We stress that the parameter π in the detection rule proposed by Xie and Siegmund [14], $\check{S}_b(\pi)$, *cannot* be interpreted as the proportion of affected sensors. Indeed, if we set $\pi = 1$ in (26) we recover the unweighted GLR-CUSUM in the case of complete uncertainty, that is the detection rule obtained by setting $L = K$ and $p = 1$ in (23). On the other hand, if we set $\pi = 1$ in $\hat{S}_b(\pi)$, defined in (24), we recover the optimal CUSUM rule in the case that all sensors are affected, which is exactly what one would expect if π is to be interpreted as the proportion of affected sensors.

3.3 Scalable schemes

We close this section by highlighting the connection between the GLR-CUSUM and the SUM-CUSUM proposed by Mei in [10]. Recall that the detection statistic of the *unweighted GLR-CUSUM* in the case of complete ignorance, that is the detection rule obtained by setting $L = K$ and $p = 1$ in (23), is

$$\max_{0 \leq s \leq t} \sum_{k=1}^K \left(Z_t^k - Z_s^k \right)^+.$$

If we interchange max and sum in this statistic, we obtain the sum of the local CUSUM statistics, $\sum_{k=1}^K Y_t^k$. This is the detection statistic of SUM-CUSUM, which was shown in [10] to be *first-order*

asymptotically optimal with respect to $\overline{\mathcal{P}}_K$, i.e., in the case of complete ignorance. The main advantage of this rule is that it is much easier to implement than the corresponding GLR-CUSUM and mixture-based CUSUM rules, as it requires running only K recursions. However, it is reasonable to expect that SUM-CUSUM should be less efficient, since it is only first-order asymptotically optimal. This intuition will be corroborated by the results of a simulation study in Section 5.

Remark: If we interchange, in a similar way, max and product in the mixture-based CUSUM rule (24), we obtain

$$M_b(\pi) := \inf \left\{ t \in \mathbb{N} : \sum_{k=1}^K \log \left(1 - \pi + \pi \exp\{Y_t^k\} \right) \geq b \right\}.$$

This is another scalable rule comparable with SUM-CUSUM, however the parameter π can no longer be interpreted as proportion of affected sensors.

Remark: From Mei [11] it follows that the detection rule

$$M_b(L) = \inf \left\{ t \in \mathbb{N} : \sum_{k=1}^L Y_t^{(k)} \geq b \right\}, \quad (27)$$

where $Y_t^{(1)} \geq Y_t^{(2)} \geq \dots \geq Y_t^{(K)}$, is *first-order* asymptotically optimal with respect to $\overline{\mathcal{P}}_L$ (at most L sensors affected). This rule should be compared with the corresponding GLR-CUSUM and mixture-based CUSUM that correspond to class $\overline{\mathcal{P}}_L$.

4 Multichart CUSUM

In this section we remain in the change-detection problem (17), where the change affects only a subset of sensors. Moreover, we assume that the various sensors are independent, i.e., (18) holds. Our focus will be on the GLR-CUSUM, defined in (10), in the special case that the change can affect exactly one sensor ($\mathcal{P} = \mathcal{P}_1$), where it takes the form

$$S = \min_{1 \leq k \leq K} S_{b_k}^k, \quad (28)$$

each b_k being a positive threshold. The detection rule (28) is also known as *multichart* CUSUM in the literature of statistical quality control and has been well studied (see, e.g., [1, 2]), especially in the case that all thresholds are equal. However, unless the sensors are homogeneous in the sense that $h_k = h$ and $g_k = g$, it is intuitively clear that it should be preferable to have unequal thresholds. The results of Section 2 imply that there is a large family of non-identical thresholds for which the multichart CUSUM is second-order asymptotically optimal. Specifically, if we set

$$b_k = b - \log p_k, \quad 1 \leq k \leq K, \quad (29)$$

where b is determined by the false alarm constraint and the p_k 's are arbitrary positive constants that add up to 1 and do not depend on γ , then from Theorem 1 it follows the resulting detection

rule is second-order asymptotically optimal for any selection of the p_k 's. Our first goal in this section is to show that there are reasonable, alternative threshold specifications with which the multichart CUSUM may fail to be even first-order asymptotically optimal. The second goal is to compare various choices for the p_k 's when the thresholds are selected according to (29).

In order to establish these results, we will rely on some well known facts about the performance characteristics of the multichart CUSUM. Thus, let $T_b^k = \inf \{t \geq 0 : Z_t^k \geq b\}$ and let η_b^k be the corresponding overshoot, i.e., $\eta_b^k := Z_{T_b^k}^k - b$. Since Z^k is a random walk with a finite second moment, the limiting expected overshoot is well defined and we have

$$\rho_k := \lim_{b \rightarrow \infty} \mathbb{E}_0^k[\eta_b^k].$$

Moreover, we have the following representation (see [25, p.32] or [9, p.44]) for the expected infimum of Z^k under \mathbb{P}_0^k :

$$\beta_k := \mathbb{E}_0^k \left[\inf_{t \in \mathbb{N}} Z_t^k \right] = \frac{\mathbb{E}_0^k[(Z_1^k)^2]}{2I_k} - \rho_k.$$

Finally, we introduce the Laplace transform of the limiting distribution of the overshoot under \mathbb{P}_0^k :

$$\delta_k := \lim_{b \rightarrow \infty} \mathbb{E}_0^k[e^{-\eta_b^k}].$$

Then, it is well known (see, e.g., [9, p.471]) that as $b_k \rightarrow \infty$

$$\mathcal{J}_k[S] = \mathbb{E}_0^k[S] \simeq \frac{b_k + \rho_k + \beta_k}{I_k}, \quad (30)$$

where $x \simeq y$ means $x - y = o(1)$. On the other hand, under \mathbb{P}_∞ we have (see, e.g., [26], [9, p.467])

$$\mathbb{E}_\infty[S] \sim \frac{1}{\sum_{k=1}^K e^{-b_k} I_k \delta_k^2}. \quad (31)$$

From (30) it is clear that if the b_k 's are chosen proportionally to the Kullback-Leibler information numbers, i.e., $b_k \propto I_k$, then the first-order asymptotic performance of the multichart CUSUM under the various scenarios is equalized, i.e., $\mathcal{J}_k[S] \sim \mathcal{J}_m[S]$ for every $1 \leq k \neq m \leq K$. However, as we show in the following proposition, with this threshold specification the multichart CUSUM loses even its first-order asymptotic optimality property with respect to \mathcal{P}_1 , unless the I_k 's are identical.

Proposition 3. *Suppose that $b_k = c_\gamma I_k$ for every $1 \leq k \leq K$, where c_γ is a constant that does not depend on k and is chosen so that the false alarm constraint be satisfied with equality, i.e., $\mathbb{E}_\infty[S] = \gamma$. Then, for every $1 \leq k \leq K$ we have:*

$$\lim_{\gamma \rightarrow \infty} \frac{\mathcal{J}_k[S]}{\inf_{T \in \mathcal{C}_\gamma} \mathcal{J}_k[T]} = \frac{I_k}{\min_{1 \leq j \leq K} I_j}.$$

Proof. Setting $b_k = c_\gamma I_k$ in (30) and (31) we obtain that $\mathcal{J}_k[S] = c_\gamma + \mathcal{O}(1)$, where $\mathcal{O}(1)$ is an asymptotically bounded term as $\gamma \rightarrow \infty$, and

$$\gamma = \mathbb{E}_\infty[S] \sim \frac{1}{\sum_{k=1}^K e^{-c_\gamma I_k} I_k \delta_k^2}. \quad (32)$$

Let $I^* := \min_{1 \leq j \leq K} I_j$. Then (32) becomes

$$\gamma \sim \frac{e^{c_\gamma I^*}}{\sum_{k: I_k = I^*} I_k \delta_k^2 + \sum_{k: I_k > I^*} e^{-c_\gamma (I_k - I^*)} I_k \delta_k^2},$$

which implies that $c_\gamma = (\log \gamma)/I^* + \mathcal{O}(1)$ and completes the proof. \square

Let us now focus on specification (29) for the thresholds. While we have seen in Theorem 1 that setting $b = \log \gamma$ guarantees $S \in \mathcal{C}_\gamma$, it is clear that this choice can be very conservative, as it does not take into account the particular pre/post-change distributions. The asymptotic expression (31) can be used to provide a much more accurate approximation for $E_\infty[S]$, even though it may not guarantee that $E_\infty[S] \geq \gamma$. Specifically, from (30) and (31) it follows that if we select threshold b in (29) as

$$b = \log \gamma + \log \left(\sum_{k=1}^K p_k I_k \delta_k^2 \right), \quad (33)$$

then $E_\infty[S] \sim \gamma$ and

$$\mathcal{J}_k[S] \simeq \frac{\log \gamma + \log \left(\sum_{j=1}^K p_j I_j \delta_j^2 \right) - \log p_k + \rho_k + \beta_k}{I_k}. \quad (34)$$

We can then use this high-order approximation in order to obtain an intuitive selection of the p_k 's. Indeed, such a selection can be obtained if we equalize, up to a first order, the asymptotic *relative* performance loss under each scenario. To be more specific, let us set

$$\bar{\mathcal{J}}_k[S] := \frac{\mathcal{J}_k[S] - \mathcal{J}_k[S^k]}{\mathcal{J}_k[S^k]}, \quad 1 \leq k \leq K,$$

where S^k is the optimal CUSUM test for detecting a change in sensor k . Assuming that the thresholds for the multichart CUSUM are selected according to (29) and (33) and that S^k is also designed so that $E_\infty[S^k] \sim \gamma$, then from (34) we obtain

$$\bar{\mathcal{J}}_k[S] \simeq \frac{C_k(p)}{\log \gamma + \rho_k + \beta_k + \log(I_k \delta_k^2)^2}, \quad (35)$$

and consequently

$$\bar{\mathcal{J}}_k[S] \sim \frac{C_k(p)}{\log \gamma},$$

where $C_k(p)$ is defined as follows:

$$C_k(p) := \log \left(\frac{\sum_{j=1}^K p_j I_j \delta_j^2}{p_k I_k \delta_k^2} \right).$$

This reveals that $\bar{\mathcal{J}}_k[S] \sim \bar{\mathcal{J}}_m[S]$ for every $1 \leq k \neq m \leq K$ when

$$p_k \propto (I_k \delta_k^2)^{-1}. \quad (36)$$

In fact, the latter specification was proposed in [24, p.223] on the basis of the observation that it guarantees $E_\infty[S_{b_k}^k] \sim E_\infty[S_{b_m}^m]$, for every $1 \leq k \neq m \leq K$.

An alternative specification for the p_k 's has been proposed in [9, p.467], according to which thresholds are selected so that $I_k \mathcal{J}_k[S] \simeq I_m \mathcal{J}_m[S]$, for every $1 \leq k \neq m \leq K$. From (34) it is clear that this is the case when

$$p_k \propto e^{\rho_k + \beta_k}.$$

Finally, if we want to have larger thresholds for sensors with larger Kullback-Leibler information numbers, which is the underlying logic in the threshold specification of Proposition 3, we can set

$$p_k \propto 1/I_k.$$

In the remainder of this section we compare the effect that the above mentioned specifications have on the asymptotic relative performance loss of a multichart CUSUM whose thresholds are selected according to (29) and (33). Specifically, we assume that $h_k = \mathcal{N}(0, 1)$ and $g_k = \mathcal{N}(\theta_k, 1)$, where $\theta_k \neq 0$. We let $K = 2$, $\gamma = 10^3$ and we compute $\bar{\mathcal{J}}_1[S]$ and $\bar{\mathcal{J}}_2[S]$, based on the asymptotic approximation (35), fixing the post-change mean in the first sensor at $\theta_1 = 1$, and letting the post-change mean in the second sensor, θ_2 , vary. In this way, we examine the relative performance loss as the relative magnitude of the anticipated change in the two sensors varies. In the Gaussian case we have the following, easily computable expressions (see, e.g., [25, p.32]) for the renewal-theoretic quantities that are included in this asymptotic approximation:

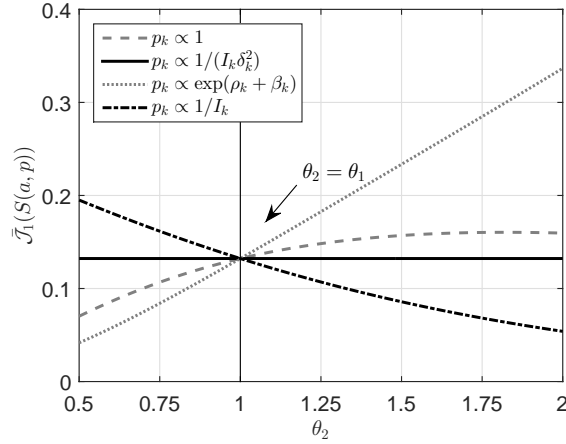
$$\begin{aligned} \beta_k &= - \sum_{n=1}^{\infty} \left[\frac{\theta_k}{\sqrt{n}} \phi(c_{n,k}) - \frac{\theta_k^2}{2} \Phi(-c_{n,k}) \right], \\ \delta_k &= \frac{2}{\theta_k^2} \exp \left\{ \sum_{n=1}^{\infty} \frac{-2}{n} \Phi(-c_{n,k}) \right\}, \end{aligned}$$

where $c_{n,k} := \theta_k \sqrt{n}/2$. The results of this computation are summarized in Fig. 1, where we see that selecting the p_k 's according to (36) leads to a more robust behavior in comparison to the other specifications when θ_2 differs significantly from θ_1 . Of course, all specifications approach the behavior of identical thresholds when θ_2 is close to θ_1 .

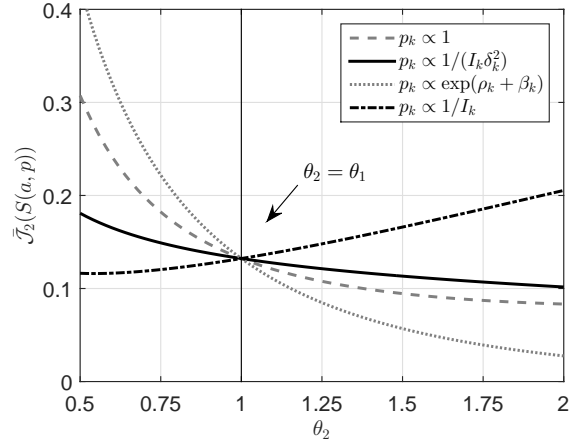
5 A simulation study

5.1 Description

In this section we present the results of a simulation study whose main goal is to compare the GLR-CUSUM, S , given by (11), the mixture-based CUSUM rules, \tilde{S} and \hat{S} , given by (14) and (24), respectively, and the SUM-CUSUM, M , defined in (27). We set $K = 5$ and $h_k = \mathcal{N}(0, 1)$, $g_k = \mathcal{N}(1, 1)$. That is, all sensors initially observe iid Gaussian observations with variance 1, and at the time of the change the mean changes from 0 to 1 in an unknown subset of these sensors. For our comparisons to be fair, we need to guarantee that all detection rules have access to the same



(a) The change occurs in the *first* sensor.



(b) The change occurs in the *second* sensor.

Figure 1: The horizontal axis in both graphs represents θ_2 , the post-change mean in the second sensor when it is the one affected by the change. The post-change mean in the first sensor, θ_1 , when this is the one affected by the change, is fixed to 1. The vertical axis in (a) (resp. (b)) represents the asymptotic relative performance loss of the multichart CUSUM when the first (resp. second) sensor is affected.

amount of prior information. We consider the following regimes: (i) no prior information ($\mathcal{P} = \overline{\mathcal{P}}_K$); (ii) knowing that *at most* L sensors can be affected ($\mathcal{P} = \overline{\mathcal{P}}_L$), where $1 \leq L \leq K$.

For the implementation of S and \tilde{S} we follow the *first* approach described in Subsection 2.4. For the implementation of \hat{S} we use the *second* approach, together with the adaptive-window based on regeneration times (15).

For all detection rules we have considered in this work, the worst case scenario is when the change occurs at time 0. Thus, when the actual affected subset is \mathcal{A} , in order to compute the worst-case detection delay of each rule we simply need to simulate it under $\mathbf{P}_0^{\mathcal{A}}$.

In order to have a direct comparison of the various detection rules and also to illustrate the second-order asymptotic optimality property, we plot the *additional* number of observations required by each rule in order to detect the change relative to the optimal CUSUM test for which the affected subset is known in advance. Thus, if T represents a generic detection rule and \mathcal{A} the true subset of affected sensors, we plot $\mathbf{E}_0^{\mathcal{A}}[T] - \mathbf{E}_0^{\mathcal{A}}[S^{\mathcal{A}}]$ against $\text{ARL}[T] := \mathbf{E}_{\infty}[T]$ for various threshold choices. In order to illustrate the first-order asymptotic optimality property, we plot the ratio $\mathbf{E}_0^{\mathcal{A}}[T] / \mathbf{E}_0^{\mathcal{A}}[S^{\mathcal{A}}]$ against $\text{ARL}[T] = \mathbf{E}_{\infty}[T]$ again for various threshold choices. In Table 1, we present numerical results for all detection rules when we choose their thresholds so that their target for the expected time to false alarm is $\gamma = 10^5$. Standard errors are given based on 50,000 Monte Carlo simulation runs.

5.2 Results

5.2.1 No prior information

We present the results that correspond to the case of no prior information in Fig. 2–Fig. 3, illustrating the notion of first-order and second-order asymptotic optimality with respect to class $\overline{\mathcal{P}}_K$, respectively. We see that when 2 sensors are affected, the mixture-based CUSUM procedures are slightly worse than the GLR-CUSUM, whereas when 4 sensors are affected, the mixture CUSUMs take the lead. In both cases the proposed schemes perform much better than SUM-CUSUM, $M(5)$, whose inflicted performance loss (relative to the optimal performance) increases much faster as γ increases. Nevertheless, its ratio over the optimal performance decreases, which supports the result that SUM-CUSUM is *first-order* asymptotically optimal.

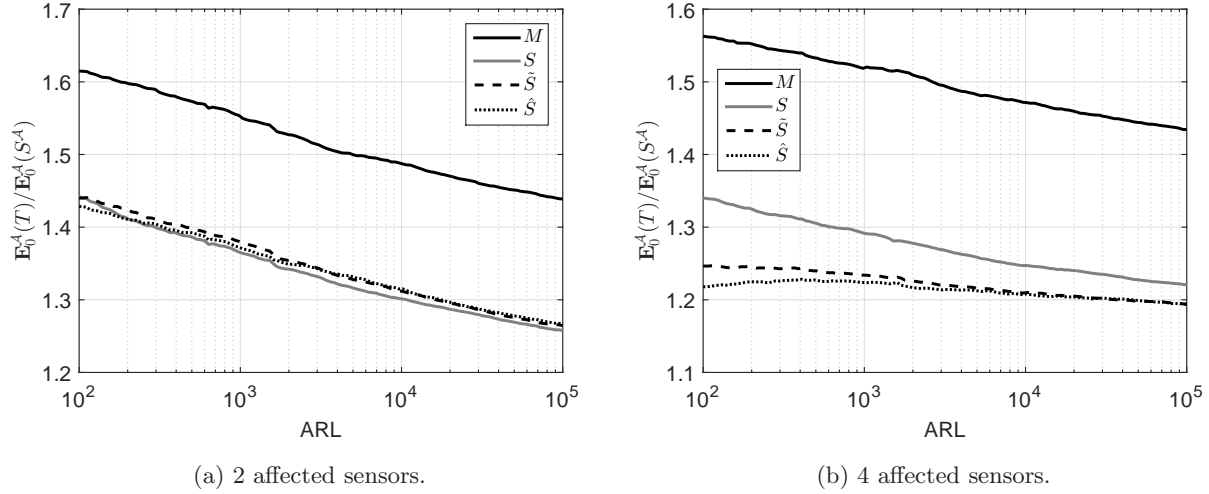


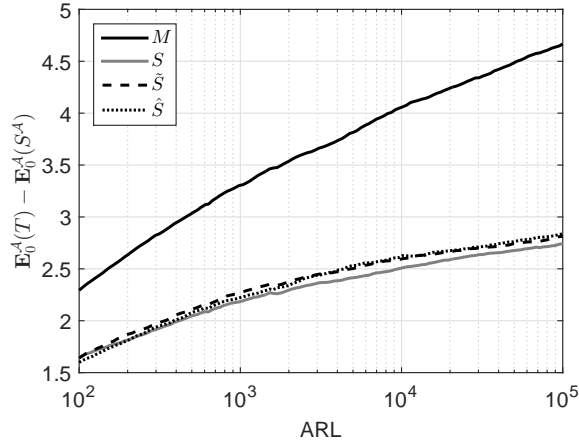
Figure 2: First-order asymptotic optimality is illustrated in the case of no prior information. The performance ratio (vertical axis) of each rule is plotted against the expected time for a false alarm (horizontal axis).

5.2.2 An upper bound on the number of affected sensors is known

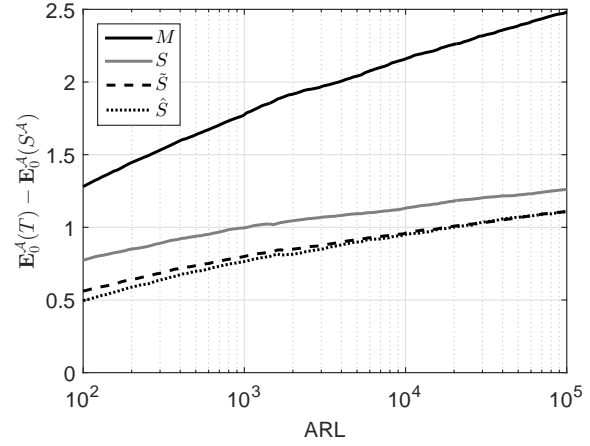
In Fig. 4–5 we assume that we know in advance that *at most* L sensors can be affected and we compare the GLR-CUSUM procedure, S , the mixture-based CUSUM procedures, \tilde{S} and $\hat{S}(L/(2K))$, and the scalable scheme $M(L)$. Again, we see that $M(L)$ performs uniformly worse than the proposed schemes in all cases.

6 Conclusions

We considered a generalized multisensor sequential change detection problem, in which a number of possibly correlated sensors are monitored on line, a global parameter vector determines their joint distribution and there is a change in an unknown subset of the components of this parameter vector.

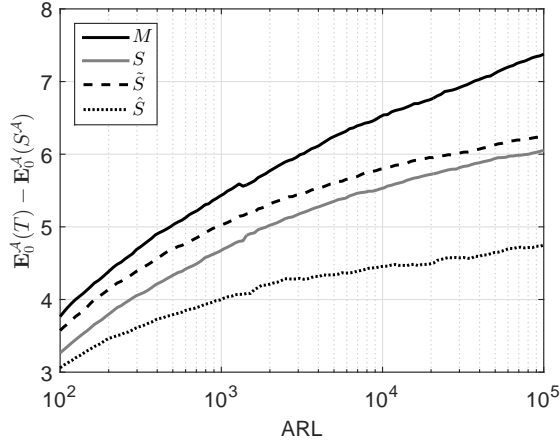


(a) 2 affected sensors.

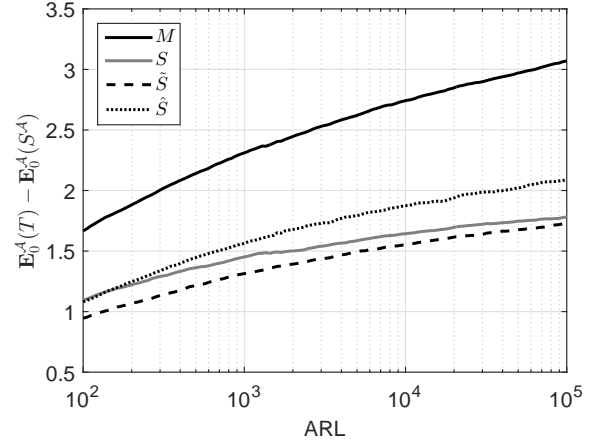


(b) 4 affected sensors.

Figure 3: Second-order asymptotic optimality is illustrated in the case of no prior information. The performance loss (vertical axis) of each rule is plotted against the expected time for a false alarm (horizontal axis).



(a) 1 sensor actually affected



(b) 3 sensors actually affected

Figure 4: Second-order asymptotic optimality illustrated when it is known that *at most 3 sensors may be affected*. The performance loss (vertical axis) of each rule is plotted against the expected time for a false alarm (horizontal axis).

In this context, we established a strong asymptotic optimality property for various CUSUM-based detection rules. In the special case that the sensors are independent and only a subset of them is affected by the change, we proposed feasible versions of the above procedures. We showed using simulation experiments that the proposed detection rules always outperform scalable detection rules, such as the one proposed in [10], and we argued that this phenomenon can be explained theoretically by the fact that the latter scheme enjoys a weaker form of asymptotic optimality. Moreover, we proposed a modification of a detection rule in [14] that is able to incorporate prior

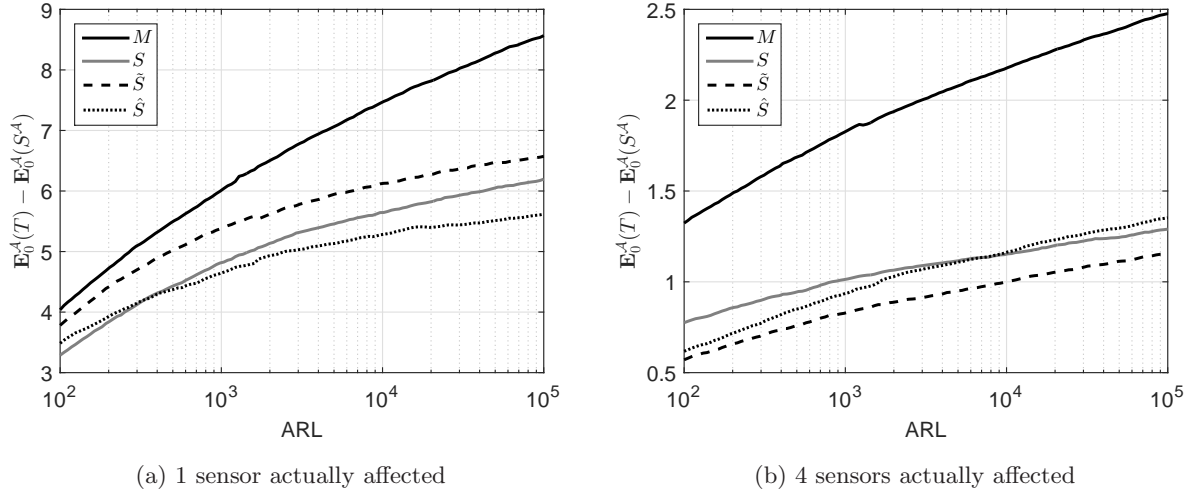


Figure 5: Second-order asymptotic optimality illustrated when it is known that *at most 4 sensors may be affected*. The performance loss (vertical axis) of each rule is plotted against the expected time for a false alarm (horizontal axis)

information. Finally, we considered the design of the multichart CUSUM in the special case that the change is known to affect exactly one sensor.

7 Acknowledgments

We would like to thank Drs. Alexander Tartakovsky, George Moustakides, Venu Veeravalli and Emmanuel Yashchin for stimulating discussions and helpful suggestions, as well as the two referees whose comments helped us improve significantly an earlier version of this paper. The work of the first author was supported by the US National Science Foundation under Grant CCF 1514245, as well as by a collaboration grant from the Simons Foundation.

Appendix

Proof of Theorem 1. From the definition of the three detection rules it is clear that $\tilde{S}_b \leq S_b$ and $\bar{\tilde{S}}_b \leq S_b$, therefore for any b and \mathcal{A} we have:

$$\mathcal{J}_{\mathcal{A}}[\bar{\tilde{S}}_b] \leq \mathcal{J}_{\mathcal{A}}[S_b] \quad \text{and} \quad \mathbb{E}_{\infty}[S_b] \geq \mathbb{E}_{\infty}[\bar{\tilde{S}}_b]$$

and

$$\mathcal{J}_{\mathcal{A}}[\tilde{S}_b] \leq \mathcal{J}_{\mathcal{A}}[S_b] \quad \text{and} \quad \mathbb{E}_{\infty}[S_b] \geq \mathbb{E}_{\infty}[\tilde{S}_b].$$

Consequently, in order to prove (i), it suffices to show that for any $b > 0$

$$\mathbb{E}_{\infty}[\bar{\tilde{S}}_b] \geq e^b \quad \text{and} \quad \mathbb{E}_{\infty}[\tilde{S}_b] \geq e^b, \quad (37)$$

Table 1: Performance of various procedures when $|\mathcal{A}|$ sensors are affected. The thresholds are selected so that the expected time to false alarm is approximately $\gamma = 10^5$.

	$ \mathcal{A} $	b	$E_\infty(T)$	SE	$E_0^A(T)$	SE
S_b^A	2	9.88	100090	450	10.64	0.02
$M_b(5)$	2	17.1	100010	455	15.30	0.03
$M_b(2)$	2	14.2	100065	450	14.21	0.03
$\hat{S}_b(0.5)$	2	9.85	100065	450	13.47	0.03
$\hat{S}_b(0.2)$	2	9.35	100105	450	13.57	0.03
$\tilde{S}_b(\overline{\mathcal{P}}_5)$	2	9.91	100105	450	13.45	0.03
$\tilde{S}_b(\overline{\mathcal{P}}_2)$	2	9.86	100115	465	13.12	0.03
$S_b(\overline{\mathcal{P}}_5)$	2	9.58	100005	445	13.38	0.03
$S_b(\overline{\mathcal{P}}_2)$	2	9.78	100005	460	13.15	0.03
S_b^A	3	9.94	100065	450	7.369	0.02
$M_b(5)$	3	17.1	100010	455	10.59	0.02
$M_b(3)$	3	15.9	100025	450	10.44	0.02
$\hat{S}_b(0.5)$	3	9.85	100065	450	9.040	0.02
$\hat{S}_b(0.3)$	3	9.63	100075	460	9.458	0.02
$\tilde{S}_b(\overline{\mathcal{P}}_5)$	3	9.91	100105	450	9.054	0.02
$\tilde{S}_b(\overline{\mathcal{P}}_3)$	3	9.90	100115	450	9.098	0.02
$S_b(\overline{\mathcal{P}}_5)$	3	9.58	100005	445	9.136	0.02
$S_b(\overline{\mathcal{P}}_3)$	3	9.67	100060	450	9.150	0.02
S_b^A	4	9.93	100010	450	5.716	0.02
$M_b(5)$	4	17.1	100010	455	8.197	0.02
$M_b(4)$	4	16.8	100050	450	8.192	0.02
$\hat{S}_b(0.5)$	4	9.85	100065	450	6.821	0.02
$\hat{S}_b(0.4)$	4	9.75	100050	450	7.068	0.02
$\tilde{S}_b(\overline{\mathcal{P}}_5)$	4	9.91	100105	450	6.826	0.02
$\tilde{S}_b(\overline{\mathcal{P}}_4)$	4	9.91	100070	450	6.870	0.02
$S_b(\overline{\mathcal{P}}_5)$	4	9.58	100005	445	6.977	0.02
$S_b(\overline{\mathcal{P}}_4)$	4	9.60	100010	450	7.006	0.02

and in order to establish (ii), it suffices to show that as $b \rightarrow \infty$

$$\mathcal{I}_A[S_b] \leq \frac{b}{I_A} + \mathcal{O}(1). \quad (38)$$

(i) In order to prove the first inequality in (37), we introduce the mixture probability measure

$$\bar{\mathbf{P}} := \sum_{A \in \mathcal{P}} p_A \mathbf{P}_0^A,$$

and denote by $\bar{\mathbf{E}}$ the corresponding expectation. The mixture-based CUSUM rule \bar{S}_b results from

repeated application, in the spirit of Lorden's [8] construction, of the one-sided sequential test:

$$\bar{T}_b := \inf \left\{ t \geq 0 : \bar{\Lambda}_t \geq e^b \right\},$$

where $\bar{\Lambda}$ is the likelihood ratio process of $\bar{\mathbf{P}}$ versus \mathbf{P}_∞ , i.e.,

$$\bar{\Lambda}_t := \frac{d\bar{\mathbf{P}}}{d\mathbf{P}_\infty}(\mathcal{F}_t) = \sum_{\mathcal{A} \in \mathcal{P}} p_{\mathcal{A}} \Lambda_t^{\mathcal{A}}, \quad t \in \mathbb{N}.$$

Consequently, from [8, Th. 2] it follows that for every $b > 0$ we have

$$\mathbf{E}_\infty[\bar{S}_b] \geq \frac{1}{\mathbf{P}_\infty(\bar{T}_b < \infty)},$$

and from Wald's likelihood ratio identity we obtain

$$\mathbf{P}_\infty(\bar{T}_b < \infty) = \bar{\mathbf{E}} \left[\bar{\Lambda}_{\bar{T}_b}^{-1} \mathbb{1}_{\{\bar{T}_b < \infty\}} \right] \leq e^{-b},$$

which completes the proof of the first inequality in (37). In order to establish the second inequality, we introduce for each subset $\mathcal{A} \in \mathcal{P}$ the corresponding Shiryaev-Roberts statistic

$$R_t^{\mathcal{A}} := \sum_{s=0}^{t-1} \exp(Z_t^{\mathcal{A}} - Z_s^{\mathcal{A}}), \quad R_0^{\mathcal{A}} := 0$$

and observe that

$$R_t^{\mathcal{A}} \geq \max_{0 \leq s < t} \exp(Z_t^{\mathcal{A}} - Z_s^{\mathcal{A}}) = \exp(\tilde{Y}_t^{\mathcal{A}}).$$

We also introduce the mixture Shiryaev-Roberts statistic $R_t := \sum_{\mathcal{A} \in \mathcal{P}} p_{\mathcal{A}} R_t^{\mathcal{A}}$ and the corresponding stopping time $\mathcal{R}_b := \inf \{ t \geq 0 : R_t \geq e^b \}$. Then, for any $b > 0$ we obtain

$$\begin{aligned} \tilde{S}_b &= \inf \left\{ t \in \mathbb{N} : \sum_{\mathcal{A} \in \mathcal{P}} p_{\mathcal{A}} \exp(\tilde{Y}_t^{\mathcal{A}}) \geq e^b \right\} \\ &\geq \inf \left\{ t \in \mathbb{N} : \sum_{\mathcal{A} \in \mathcal{P}} p_{\mathcal{A}} R_t^{\mathcal{A}} \geq e^b \right\} = \mathcal{R}_b, \end{aligned}$$

thus, it suffices to show that $\mathbf{E}_\infty[\mathcal{R}_b] \geq e^b$. In order to see this, note that for any b and t we have $0 \leq R_t < e^b$ on the event $\{\mathcal{R}_b > t\}$, therefore

$$\begin{aligned} \int_{\{\mathcal{R}_b > t\}} |R_t - t| d\mathbf{P}_\infty &\leq \int_{\{\mathcal{R}_b > t\}} R_t d\mathbf{P}_\infty + \int_{\{\mathcal{R}_b > t\}} t d\mathbf{P}_\infty \\ &\leq e^b \mathbf{P}_\infty(\mathcal{R}_b > t) + \int_{\{\mathcal{R}_b > t\}} \mathcal{R}_b d\mathbf{P}_\infty. \end{aligned}$$

For any given b , the upper bound goes to 0 as $t \rightarrow \infty$ as long as $\mathbf{E}_\infty[\mathcal{R}_b] < \infty$. But this can be assumed without any loss of generality, since otherwise the desired inequality holds trivially. Therefore, we can apply the optional sampling theorem to the \mathbf{P}_∞ -martingale $\{R_t - t\}_{t \in \mathbb{N}}$ and

the stopping time \mathcal{R}_b and obtain $\mathbb{E}_\infty[\mathcal{R}_b] = \mathbb{E}_\infty[R_{\mathcal{R}_b}] \geq e^b$, where the inequality holds from the definition of the stopping rule \mathcal{R}_b . This completes the proof of the second inequality in (37).

(ii) For any $b > 0$ and $\mathcal{A} \in \mathcal{P}$, from the definition of S_b in (11) it follows that $S_b \leq S_{b-\log p_{\mathcal{A}}}^{\mathcal{A}}$, where $S_{b-\log p_{\mathcal{A}}}^{\mathcal{A}}$ is the CUSUM rule defined in (5) with threshold $b-\log p_{\mathcal{A}}$. Consequently, as $b \rightarrow \infty$ we have

$$\mathcal{J}_{\mathcal{A}}[S_b] \leq \frac{b - \log p_{\mathcal{A}}}{I_{\mathcal{A}}} + \mathcal{O}(1) = \frac{b}{I_{\mathcal{A}}} + \mathcal{O}(1),$$

which completes the proof of (38). \square

Proof of Proposition 1. Fix $b > 0$ and $\mathcal{A} \in \mathcal{P}$. (i) We will first prove the result for the GLR-CUSUM, S_b . On the event $\{S_b > \nu\}$ we have

$$S_b = \inf \left\{ t > \nu : \max_{\mathcal{B} \in \mathcal{P}} (Y_t^{\mathcal{B}} - Y_\nu^{\mathcal{B}} + Y_\nu^{\mathcal{B}} + \log p_{\mathcal{B}}) \geq b \right\}.$$

For each $\mathcal{B} \in \mathcal{P}$, $\{Y_t^{\mathcal{B}} - Y_\nu^{\mathcal{B}}\}_{t \geq \nu}$ depends only on the post-change observations. Therefore, the detection statistic depends on the pre-change observations only through the non-negative random variables $Y_\nu^{\mathcal{B}}$, $\mathcal{B} \in \mathcal{P}$. Moreover, the detection statistic is increasing in these quantities, which means that, for any possible change point ν , the worst-possible scenario for the history of observations up to ν is that $Y_\nu^{\mathcal{B}} = 0$ for every $\mathcal{B} \in \mathcal{P}$:

$$\text{ess sup } \mathbb{E}_\nu^{\mathcal{A}}[(S_b - \nu)^+ | \mathcal{F}_\nu] = \mathbb{E}_\nu^{\mathcal{A}}[(S_b - \nu)^+ | Y_\nu^{\mathcal{B}} = 0, \forall \mathcal{B} \in \mathcal{P}].$$

However, when $Y_\nu^{\mathcal{B}} = 0$ for every $\mathcal{B} \in \mathcal{P}$, each sequence $\{Y_t^{\mathcal{B}} - Y_\nu^{\mathcal{B}}\}_{t \geq \nu}$ under $\mathbb{P}_\nu^{\mathcal{A}}$ has the same distribution as $\{Y_t^{\mathcal{B}}\}_{t \geq 0}$ under $\mathbb{P}_0^{\mathcal{A}}$. Thus, for every change-point ν we have

$$\mathbb{E}_\nu^{\mathcal{A}}[(S_b - \nu)^+ | Y_\nu^{\mathcal{B}} = 0, \forall \mathcal{B} \in \mathcal{P}] = \mathbb{E}_0^{\mathcal{A}}[S_b],$$

which proves that $\mathcal{J}_{\mathcal{A}}[S_b] = \mathbb{E}_0^{\mathcal{A}}[S_b]$.

(ii) In order to prove the corresponding result for \bar{S}_b , we note that on the event $\{\bar{S}_b > \nu\}$ we have

$$\begin{aligned} \bar{S}_b &= \inf \left\{ t > \nu : \max_{0 \leq s \leq t} \sum_{\mathcal{B} \in \mathcal{P}} p_{\mathcal{B}} \exp(Z_t^{\mathcal{B}} - Z_s^{\mathcal{B}}) \geq e^b \right\} \\ &\leq \inf \left\{ t > \nu : \max_{\nu \leq s \leq t} \sum_{\mathcal{B} \in \mathcal{P}} p_{\mathcal{B}} \exp(Z_t^{\mathcal{B}} - Z_s^{\mathcal{B}}) \geq e^b \right\} \\ &:= \bar{S}_b^{(\nu)}. \end{aligned}$$

For every change point ν , $\bar{S}_b^{(\nu)}$ under $\mathbb{P}_\nu^{\mathcal{A}}$ has the same distribution as \bar{S}_b under $\mathbb{P}_0^{\mathcal{A}}$ and is independent of \mathcal{F}_ν . Therefore,

$$\mathbb{E}_\nu^{\mathcal{A}}[(\bar{S}_b - \nu)^+ | \mathcal{F}_\nu] \leq \mathbb{E}_\nu^{\mathcal{A}}[\bar{S}_b^{(\nu)}] = \mathbb{E}_0^{\mathcal{A}}[\bar{S}_b],$$

which proves that $\mathcal{J}_A[\bar{S}_b] \leq \mathbb{E}_0^A[\bar{S}_b]$, and consequently $\mathcal{J}_A[\bar{S}_b] = \mathbb{E}_0^A[\bar{S}_b]$.

(iii) Finally, on the event $\{\tilde{S}_b > \nu\}$ we have

$$\begin{aligned}\tilde{S}_b &= \inf \left\{ t > \nu : \sum_{\mathcal{B} \in \mathcal{P}} p_{\mathcal{B}} \exp\{\tilde{Y}_t^{\mathcal{B}}\} \geq b \right\} \\ &= \inf \left\{ t > \nu : \sum_{\mathcal{B} \in \mathcal{P}} p_{\mathcal{B}} \exp\{\tilde{Y}_t^{\mathcal{B}} - \tilde{Y}_{\nu}^{\mathcal{B}} + \tilde{Y}_{\nu}^{\mathcal{B}}\} \geq b \right\}.\end{aligned}$$

Fix $\mathcal{B} \in \mathcal{P}$. When $\tilde{Y}_{\nu}^{\mathcal{B}} < 0$, then from recursion (7) it follows that $\{\tilde{Y}_t^{\mathcal{B}}\}_{t > \nu}$ is independent of \mathcal{F}_{ν} and that its distribution under \mathbb{P}_{ν}^A is the same as that of $\{\tilde{Y}_t^{\mathcal{B}}\}_{t > 0}$ under \mathbb{P}_0^A . When $\tilde{Y}_{\nu}^{\mathcal{B}} \geq 0$, a similar argument as in (i) shows that the worst case is $\tilde{Y}_{\nu}^{\mathcal{B}} = 0$. Thus, we conclude that $\mathcal{J}_A[\tilde{S}_b] = \mathbb{E}_0^A[\tilde{S}_b]$. \square

Proof of Proposition 2. Recall the notation $C(\mathcal{P}) := \sum_{\mathcal{B} \in \mathcal{P}} p^{|\mathcal{B}|}$.

(i) We observe that $\log C(\mathcal{P}_L) = L \log p + \log |\mathcal{P}_L|$ and

$$\max_{\mathcal{A} \in \mathcal{P}_L} \sum_{k \in \mathcal{A}} \left(Z_t^k - Z_s^k + \log p \right) = \sum_{k=1}^L \left(Z_{s:t}^{(k)} + \log p \right),$$

which proves (22). Moreover, we observe that $\log C(\bar{\mathcal{P}}_L) = \log \left(\sum_{k=1}^L |\mathcal{P}_k| p^k \right)$ and that

$$\max_{\mathcal{A} \in \mathcal{P}} \sum_{k \in \mathcal{A}} \left(Z_t^k - Z_s^k + \log p \right) = \sum_{k=1}^L \left(Z_{s:t}^{(k)} + \log p \right)^+$$

whenever there is at least one k such that $Z_t^k - Z_s^k + \log p > 0$. On the other hand, if $Z_t^k - Z_s^k + \log p \leq 0$ for every $1 \leq k \leq K$, the detection statistic cannot reach a positive threshold, which completes the proof of (23).

(ii) We observe that for any $\pi \in (0, 1)$ and $p = \pi/(1 - \pi)$ we have

$$\begin{aligned}& \prod_{k=1}^K \left(1 - \pi + \pi \frac{\Lambda_t^k}{\Lambda_s^k} \right) \\ &= (1 - \pi)^K + \sum_{\mathcal{A} \in \bar{\mathcal{P}}_K} (1 - \pi)^{K - |\mathcal{A}|} \pi^{|\mathcal{A}|} \prod_{k \in \mathcal{A}} \frac{\Lambda_t^k}{\Lambda_s^k} \\ &= (1 - \pi)^K \left[1 + \sum_{\mathcal{A} \in \bar{\mathcal{P}}_K} p^{|\mathcal{A}|} \frac{\Lambda_t^{\mathcal{A}}}{\Lambda_s^{\mathcal{A}}} \right],\end{aligned}$$

which proves (24). \square

Proof of Theorem 2. (i) Fix b . We need to show that

$$\mathbb{E}_{\infty}[\hat{S}_b(\pi)] \geq e^b \quad \text{and} \quad \mathbb{E}_{\infty}[\check{S}_b(\pi)] \geq \frac{e^b}{|\bar{\mathcal{P}}_K|}, \quad (39)$$

where $\pi \in (0, 1)$ in the first and $\pi \in (0, 1]$ in the second inequality. In order to prove the first inequality, we recall that $\hat{S}_b(\pi)$ can be obtained with the repeated application of the one-sided sequential test $\hat{T}_b(\pi)$, defined in (25). Therefore, from [8, Th. 2] it follows that

$$\mathbb{E}_\infty[\hat{S}_b(\pi)] \geq \frac{1}{\mathbb{P}_\infty(\hat{T}_b(\pi) < \infty)}$$

and from Wald's likelihood ratio identity we obtain

$$\mathbb{P}_\infty(\hat{T}_b(\pi) < \infty) = \hat{\mathbb{E}} \left[\exp \left\{ -\hat{Z}_{\hat{T}_b(\pi)} \right\} \mathbb{1}_{\{\hat{T}_b(\pi) < \infty\}} \right] \leq e^{-b},$$

where \hat{Z} is the logarithm of the likelihood ratio $\hat{\Lambda}^\pi$ defined in (25), i.e.,

$$\hat{Z}_t := \log \hat{\Lambda}_t = \sum_{k=1}^K \log \left(1 - \pi + \pi \Lambda_t^k \right).$$

This completes the proof of the first inequality in (39). In order to prove the second inequality, it suffices to observe that for any $b > 0$ and $\pi \in (0, 1]$ we have

$$\hat{S}_b(\pi) \geq \check{S}_b(\pi) \geq \inf \left\{ t \geq 0 : \max_{0 \leq s \leq t} \sum_{k=1}^K (Z_t^k - Z_s^k)^+ \geq b \right\}$$

and that the lower bound coincides (up to a constant) with the unweighted GLR-CUSUM in the case of complete uncertainty, that is the detection rule obtained by setting $L = K$ and $\pi = 1$ in (23).

(ii) When $\pi = 1$, $\check{S}_b(\pi)$ coincides with the GLR-CUSUM in the case of complete uncertainty, whose asymptotic optimality has already been established. Therefore, without loss of generality, we assume that $\pi \in (0, 1)$. Due to the pathwise inequality $\hat{S}_b(\pi) \geq \check{S}_b(\pi)$, it suffices to show that for every subset \mathcal{A} we have

$$\mathcal{J}_\mathcal{A}[\hat{S}_b(\pi)] \leq \frac{b}{I_\mathcal{A}} + \mathcal{O}(1),$$

where $\mathcal{O}(1)$ is a bounded term as $b \rightarrow \infty$. From [8, Th. 2] it follows that $\mathcal{J}_\mathcal{A}[\hat{S}_b(\pi)] \leq \mathbb{E}_0^\mathcal{A}[\hat{T}_b(\pi)]$ for every $b > 0$, whereas for every t we observe that

$$\begin{aligned} \hat{Z}_t &= \sum_{k=1}^K \log \left(1 - \pi + \pi \exp(Z_t^k) \right) \\ &\geq \sum_{k \in \mathcal{A}} (Z_t^k + \log \pi) + \sum_{k \notin \mathcal{A}} \log(1 - \pi) \\ &= Z_t^\mathcal{A} + c_\mathcal{A}, \end{aligned}$$

where $c_\mathcal{A} := |\mathcal{A}| \log \pi + (K - |\mathcal{A}|) \log(1 - \pi)$. Therefore, $\hat{T}_b(\pi) \leq \inf \{ t : Z_t^\mathcal{A} \geq b + c_\mathcal{A} \}$ for every $b > 0$, and consequently

$$\mathbb{E}_0^\mathcal{A}[\hat{T}_b(\pi)] \leq \frac{b + c_\mathcal{A}}{I_\mathcal{A}} + \mathcal{O}(1),$$

which completes the proof. \square

References

- [1] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blažek, and H. Kim, “A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods,” *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3372–3382, 2006.
- [2] —, “Detection of intrusions in information systems by sequential changepoint methods (with discussion),” *Statistical Methodology*, vol. 3, no. 3, pp. 252–340, 2006.
- [3] C. Lévy-Leduc and F. Roueff, “Detection and localization of change-points in high-dimensional network traffic data,” *Annals of Applied Statistics*, vol. 3, no. 2, pp. 637–662, 2009.
- [4] C. Sonesson and D. Bock, “A review and discussion of prospective statistical surveillance in public health,” *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, vol. 166, no. 1, pp. 5–21, 2003.
- [5] G. Shmueli and H. Burkom, “Statistical challenges facing early outbreak detection in bio-surveillance,” *Technometrics*, vol. 52, no. 1, pp. 39–51, February 2010.
- [6] E. S. Page, “Continuous inspection schemes,” *Biometrika*, vol. 41, no. 1, pp. 100–115, 1954.
- [7] G. V. Moustakides, “Optimal stopping times for detecting changes in distributions,” *The Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, 1986.
- [8] G. Lorden, “Procedures for reacting to a change in distribution,” *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, 1971.
- [9] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. CRC Press, 2014.
- [10] Y. Mei, “Efficient scalable schemes for monitoring a large number of data streams,” *Biometrika*, vol. 97, no. 2, pp. 419–433, 2010.
- [11] —, “Quickest detection in censoring sensor networks,” in *IEEE International Symposium on Information Theory*, 2011, pp. 2148–2152.
- [12] K. Liu, Y. Mei, and J. Shi, “An adaptive sampling strategy for online high-dimensional process monitoring,” *Technometrics*, 2015, to appear.
- [13] T. Banerjee and V. V. Veeravalli, “Data-efficient quickest outlying sequence detection in sensor networks.”
- [14] Y. Xie and D. Siegmund, “Sequential multi-sensor change-point detection,” *The Annals of Statistics*, vol. 41, no. 2, pp. 670–692, 2013.

- [15] Y. C. Chen, T. Banerjee, A. D. Dominguez-Garcia, and V. V. Veeravalli, “Quickest line outage detection and identification,” *Power Systems, IEEE Transactions on*, vol. PP, no. 99, pp. 1–10, 2015.
- [16] D. Siegmund and E. S. Venkatraman, “Using the generalized likelihood ratio statistic for sequential detection of a change-point,” *Annals of Statistics*, vol. 23, no. 1, pp. 255–271, 1995.
- [17] V. Raghavan and V. V. Veeravalli, “Quickest change detection of a markov process across a sensor array,” *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1961–1981, April 2010.
- [18] M. Ludkovski, “Bayesian quickest detection in sensor arrays,” *Sequential Analysis*, vol. 31, no. 4, pp. 481–504, 2012.
- [19] O. Hadjiliadis, H. Zhang, and H. V. Poor, “One shot schemes for decentralized quickest change detection,” *Information Theory, IEEE Transactions on*, vol. 55, no. 7, pp. 3346–3359, July 2009.
- [20] H. Zhang, N. Rodosthenous, and O. Hadjiliadis, “Robustness of the N -CUSUM stopping rule in a wiener disorder problem,” *The Annals of Applied Probability*, 2015, to appear.
- [21] I. V. Nikiforov, “A generalized change detection problem,” *Information Theory, IEEE Transactions on*, vol. 41, no. 1, pp. 171–187, January 1995.
- [22] —, “Two strategies in the problem of change detection and isolation,” *Information Theory, IEEE Transactions on*, vol. 43, no. 2, pp. 770–776, March 1997.
- [23] T. Oskiper and H. V. Poor, “Online activity detection in a multiuser environment using the matrix CUSUM algorithm,” *Information Theory, IEEE Transactions on*, vol. 48, no. 2, pp. 477–493, February 2002.
- [24] A. G. Tartakovsky, “Multidecision quickest changepoint detection: Previous achievements and open problems,” *Sequential Analysis*, vol. 27, pp. 201–231, 2008.
- [25] M. Woodroffe, *Nonlinear Renewal Theory in Sequential Analysis*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1982.
- [26] R. A. Khan, “Detecting changes in probabilities of a multi-component process,” *Sequential Analysis*, vol. 14, no. 4, pp. 375–388, 1995.
- [27] E. Yashchin, “Regenerative likelihood ratio control schemes,” in *Frontiers in Statistical Quality Control 11*, ser. Frontiers in Statistical Quality Control, S. Knoth and W. Schmid, Eds., 2015, vol. 11.
- [28] —, “Change-point models in industrial applications,” in *Proceedings of the Second World Congress of Nonlinear Analysts*, vol. 30, no. 7, December 1997, pp. 3997–4006.

- [29] —, “Likelihood ratio methods for monitoring parameters of a nested random effect model,” *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 729–738, June 1995.